# Exploring the Effects of Health Factors and Age on Heart Attack Risk

Agyeah Stephen

**Introduction**

In this project, we analyze data from 1,319 patients, focusing on key predictors such as troponin levels, glucose levels, blood pressure, and age, this report analyzes heart attack patient data to identify significant predictors of positive or negative outcomes and to model these outcomes using statistical learning techniques.The objective is to use the insights gained from the data and models to better understand the risk factors and their relationships with heart attack outcomes.

**Problem Statement**

Heart attacks remain a leading cause of mortality worldwide, often arising from clinical and demographic factors. Early detection and prevention are critical to improving patient outcomes. This project aims to build predictive models that identify the likelihood of a heart attack based on variables such as troponin levels, glucose, age, blood pressure, and other clinical variables. By analyzing these factors, the project seeks to determine which predictors are most significant in influencing outcomes and to provide good insights for decision making.

**Dataset Overview**

The dataset contains 1,319 observations with the following variables:

*Class* : Outcomes (0 = negative, 1 = positive)
*Troponin* : Troponin levels (indicator of heart muscle damage)
*Glucose* : Blood glucose levels
*Pressurehigh*: High blood pressure readings
*Pressurelow* : Low blood pressure readings
*Age* : Age of the patient
*Impulse* : Heart impulse measurement
*Gender* : Male or Female

```
glimpse(heart_attack)

## Rows: 1,319
## Columns: 9
## $ age           <dbl> 64, 21, 55, 64, 55, 58, 32, 63, 44, 67, 44, 63, 64,
54, …
## $ gender        <dbl> 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0,
0, 0,…
## $ impluse       <dbl> 66, 94, 64, 70, 64, 61, 40, 60, 60, 61, 60, 60, 60,
94, …
## $ pressurehight <dbl> 160, 98, 160, 120, 112, 112, 179, 214, 154, 160,
166, 15…
## $ pressurelow   <dbl> 83, 46, 77, 55, 65, 58, 68, 82, 81, 95, 90, 83, 99,
67, …
## $ glucose       <dbl> 160, 296, 270, 270, 300, 87, 102, 87, 135, 100, 102,
198…
## $ kcm           <dbl> 1.800, 6.750, 1.990, 13.870, 1.080, 1.830, 0.710,
300.00…
## $ troponin      <dbl> 0.012, 1.060, 0.003, 0.122, 0.003, 0.004, 0.003,
2.370, …
## $ class         <chr> "negative", "positive", "negative", "positive",
"negativ…
```

### Checking for missing data

```
anyNA(heart_attack)

## [1] FALSE
```

The function returns *FALSE*, meaning there are no missing values in the dataset heart attack . This is important because missing data can skew the results of statistical models, and having a complete dataset ensures that our analysis is accurate and reliable.

### Research Questions and Hypothesis

We aim to explore the relationship between the heart attack variables and the outcomes.

### question 1

What is the relationship between age, glucose levels, blood pressure, and troponin in patients, and how do they correlate with each other?
This question helps identify how these variables interact and whether certain combinations of variables are more indicative of heart attack outcomes than others.

### question 2

Does gender and age significant factor in determining the likelihood of a positive or negative outcome? This question examines whether gender differences play a major role in heart attack outcomes, providing insights into potential demographic disparities.

### Hypothesis

**Research Question 1**:
What is the relationship between age, glucose levels, blood pressure, and troponin in patients, and how do they correlate with each other?

**Null Hypothesis ($H_o$)**:
There is no significant correlation between age, glucose levels, blood pressure, and troponin levels in patients.

**Alternative Hypothesis ($H_a$)**:
There is a significant correlation between these clinical metrics, indicating that changes in

one metric (e.g., troponin) are associated with changes in others (e.g., glucose or blood pressure).

**Research Question 2**:
Is age a significant factor in determining the likelihood of a positive or negative outcome?

**Null Hypothesis ($H_o$)**:
Age has a significant effect on heart attack outcomes, suggesting that the likelihood of a positive or negative result may differ.

**Alternative Hypothesis ($H_a$)**:
Age has no significant effect on the likelihood of a positive or negative heart attack outcome.

**Research Question 1 Variables**:

Age : representing the patient's age.
Glucose Levels : variable measuring blood glucose.
High Blood Pressure (pressurehight) : variable representing high blood pressure.
Low Blood Pressure (pressurelow) : variable representing low blood pressure.
Troponin Levels : Continuous variable indicating the level of troponin, a key marker for heart attack severity

Main response variables are
Class : Positive or negative outcome
Troponin Levels : Continuous variable indicating the level of troponin

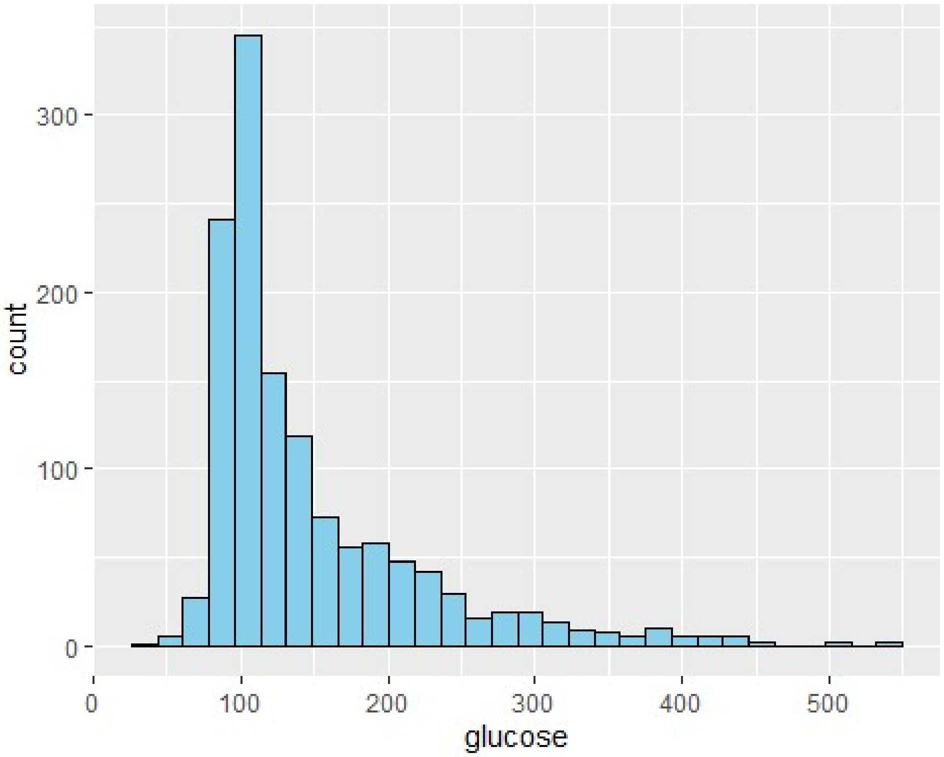
**Research Question 2 Variables**:
Age : representing the patient's age.
Troponin Levels: Continuous variable.
Glucose Levels: Continuous variable.
High Blood Pressure (pressurehight): Continuous variable.

**Main Response Variable**:
Class: Binary outcome (0 = negative outcome, 1 = positive outcome)

**Exploratory Data Analysis**:
**Glucose**
Understanding the distribution of glucose levels in the dataset is very important, as glucose levels can indicate underlying health issues that may affect heart attack outcomes. The following histogram visualizes the distribution of glucose levels across all patients

```
library(ggplot2)
heart_attack |>
ggplot(aes(glucose))+
  geom_histogram(col="black",fill="skyblue" )
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



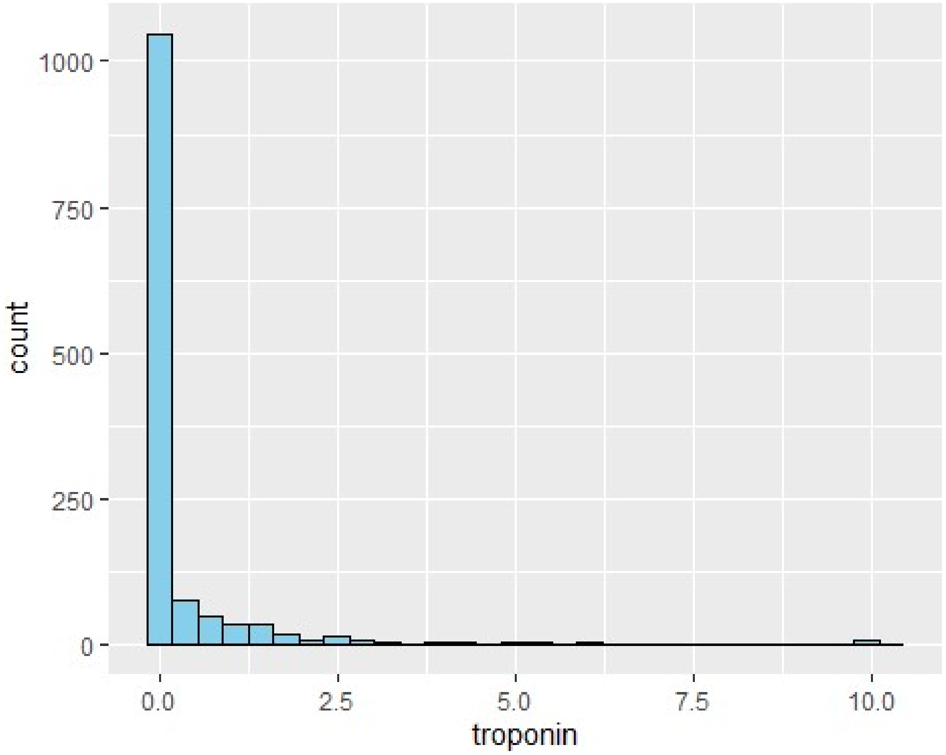
The histogram reveals that the glucose data is skewed to the right, indicating that most patients have lower glucose levels, with fewer patients showing higher values.The right skewness suggests that while the majority of glucose levels are within a normal range, there is a long tail of higher values.There are also a few outliers with exceptionally high glucose levels, which could correspond to patients with positive outcomes.

**Troponin**

Understanding the distribution of troponin levels is crucial, as troponin is a wellknown marker for heart attack severity. The histogram below illustrates the distribution of troponin levels in the dataset

```
heart_attack |>
ggplot(aes(troponin))+
  geom_histogram(col="black",fill="skyblue" )
```

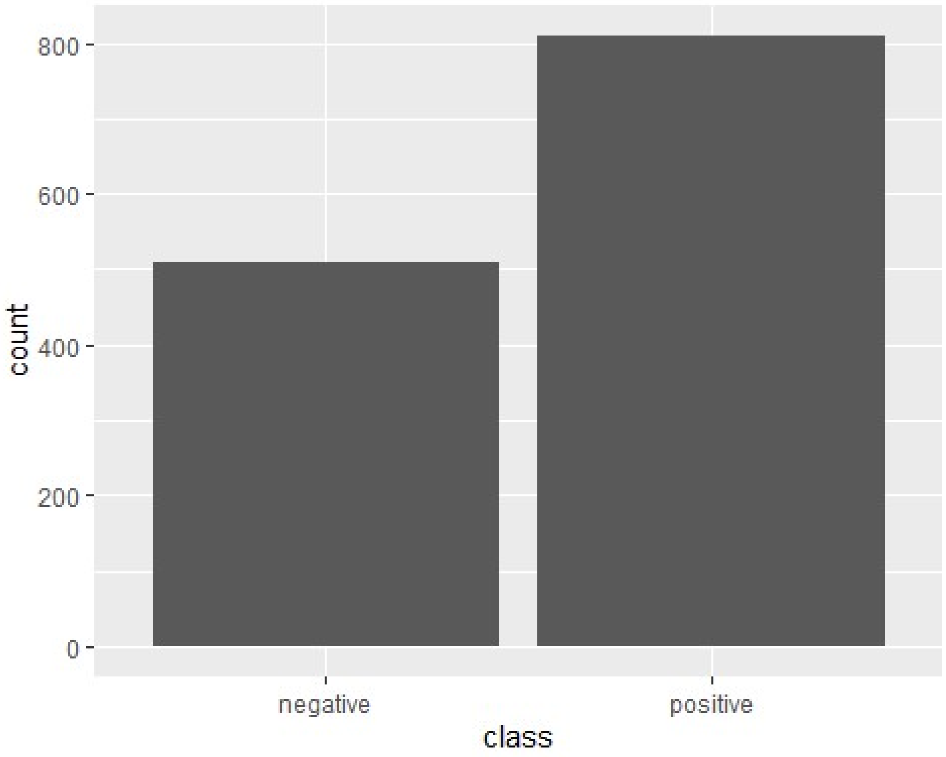
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The histogram shows a right skewed pattern just like the gluclose, where most patients have relatively low troponin levels, with a few exhibiting significantly higher values. This skewness suggests that while many individuals in the dataset have normal or slightly elevated troponin levels, there is a subset with extremely high values.The presence of outliers in the higher troponin range is very important. Elevated troponin levels typically suggest that these patients may have experienced severe heart attacks.

To help us better understand the distribution of heart attack outcomes in the dataset, we create a bar plot. This represent the count of positive (1) and negative (0) outcomes, which helps us assess the balance between the two classes.

```
heart_attack |>
ggplot(aes(class))+
  geom_bar()
```



The bar plot shows the distribution of positive and negative outcomes, it is clear that the positive outcomes are significantly higher than the negative outcomes (no heart attack),this indicates that the number of people who are positive for heart attact are more than those with a negative outcome.

**Converting Variables to Factors**:
The class (heart attack outcome) and gender variables were converted to factors to ensure they are treated correctly as categorical variables in the model

```
heart_attack <- heart_attack |>
  mutate(class = factor(class),gender=factor(gender))
heart_attack

## # A tibble: 1,319 × 9
##      age gender impluse pressurehight pressurelow glucose    kcm troponin
class
##    <dbl> <fct>    <dbl>         <dbl>       <dbl>   <dbl>  <dbl>    <dbl>
<fct>
##  1    64 1           66           160          83     160   1.8     0.012
negat…
##  2    21 1           94            98          46     296   6.75    1.06
posit…
##  3    55 1           64           160          77     270   1.99    0.003
negat…
##  4    64 1           70           120          55     270  13.9     0.122
posit…
```

```
##  5    55 1           64           112          65      300   1.08    0.003
negat…
##  6    58 0           61           112          58       87   1.83    0.004
negat…
##  7    32 0           40           179          68      102   0.71    0.003
negat…
##  8    63 1           60           214          82       87 300       2.37
posit…
##  9    44 0           60           154          81      135   2.35    0.004
negat…
## 10    67 1           61           160          95      100   2.84    0.011
negat…
## # ℹ 1,309 more rows
```

This transformation ensures that the class variable is treated as a categorical response (0 or 1) and gender as a categorical predictor (Male or Female), which is important for the logistic regression model

**Creating a New Variable**:
**Glucose Levels**
We created a new variable, glucosel_level, to classify patients based on their glucose levels;

```
heart_attack <- heart_attack |>
  mutate(glucose_level = ifelse(glucose > 140,"High","Normal"))
```

This new binary variable categorizes patients into High glucose levels (greater than 140) and Normal glucose levels.Glucose levels are an important predictor for heart attack risks and may improve the model's ability to predict outcomes. By defining a threshold for glucose_level, we simplify the variable into two distinct categories for easy interpretation and analysis.

**Data Summary: Descriptive Statistics**
To better understand the characteristics of the dataset and the distribution of key variables, we computed various summary statistics for glucose levels, age, and troponin levels. These summary statistics help provide a foundational understanding of the data, and guide the analysis for modeling heart attack outcomes.

**Summary Statistics for Glucose Levels,Troponin Levels and Age**
We first computed overall summary statistics for glucose and troponin levels to understand their distribution in the dataset

Now the Summary staticstics gives us
Mean : The average for both glucose and troponin levels across all patients.
Standard Deviation (SD) : Measures the spread of glucose and troponin levels in the dataset.
Quartiles (Q1, Median, Q3) : These give us an idea of the distribution, showing where 50% of values lie.

Min and Max : Shows the lowest and highest of both glucose and troponin levels found in the data set.

```
heart_attack |>
  summarise(Mean_glucose = mean(glucose),
            SD_glucose = sd(glucose),
            Min_glucose = min(glucose),
            Q1_glucose = quantile(glucose,0.25),
            Median_glucose = median(glucose),
            Q3_glucose = quantile(glucose,0.75),
            Max_glucose = max(glucose))
```

```
## # A tibble: 1 × 7
##   Mean_glucose SD_glucose Min_glucose Q1_glucose Median_glucose Q3_glucose
##          <dbl>      <dbl>       <dbl>      <dbl>          <dbl>      <dbl>
## 1         147.       74.9          35         98            116       170.
## # ℹ 1 more variable: Max_glucose <dbl>
```

```
heart_attack |>
  summarise(Mean_troponin = mean(troponin),
            SD_tropoin = sd(troponin),
            Min_troponin = min(troponin),
            Q1_troponin = quantile(troponin,0.25),
            Median_troponin = median(troponin),
            Q3_troponin = quantile(troponin,0.75),
            Max_troponin = max(troponin))
```

```
## # A tibble: 1 × 7
##   Mean_troponin SD_tropoin Min_troponin Q1_troponin Median_troponin
Q3_troponin
##           <dbl>      <dbl>        <dbl>       <dbl>           <dbl>
<dbl>
## 1         0.361       1.15        0.001       0.006           0.014
0.0855
## # ℹ 1 more variable: Max_troponin <dbl>
```

```
heart_attack |>
  summarise(Mean_age = mean(age),
            SD_age = sd(age),
            Min_age = min(age),
            Q1_age = quantile(age,0.25),
            Median_age = median(age),
            Q3_age = quantile(age,0.75),
            Max_age = max(age))
```

```
## # A tibble: 1 × 7
##   Mean_age SD_age Min_age Q1_age Median_age Q3_age Max_age
##      <dbl>  <dbl>   <dbl>  <dbl>      <dbl>  <dbl>   <dbl>
## 1     56.2   13.6      14     47         58     65     103
```

In addition to these overall statistics, we perform grouped analysis by dividing the data based on outcome classes (i.e., positive vs. negative heart attack outcomes) and glucose levels(High and Normal). This grouped analysis helps us to identify differences in the characteristics of patients who experienced a heart attack versus those who did not. We computed summary statistics such as the mean, standard deviation, min, max, and quartiles for each group to observe the variation and average values within each category.

### Glucose

```
heart_attack |>
  group_by(class) |>
  summarise(Mean_glucose = mean(glucose),
            SD_glucose = sd(glucose),
            Min_glucose = min(glucose),
            Q1_glucose = quantile(glucose,0.25),
            Median_glucose = median(glucose),
            Q3_glucose = quantile(glucose,0.75),
            Max_glucose = max(glucose))
```

```
## # A tibble: 2 × 8
##   class Mean_glucose SD_glucose Min_glucose Q1_glucose Median_glucose
Q3_glucose
##   <fct>        <dbl>      <dbl>       <dbl>      <dbl>          <dbl>
<dbl>
## 1 nega…         150.       78.4          60         98            117
184
## 2 posi…         145.       72.6          35         98            116
166
## # ℹ 1 more variable: Max_glucose <dbl>
```

Based on the summary data provided above for the two classes (negative and positive), we can observe that the Negative class have higher glucose levels on average, with more variability in glucose levels and the Positive Class has slightly lower glucose levels on average, but with a larger proportion of individuals having low glucose values. This suggests that glucose level can be a significant factor in differentiating between the negative and positive classes.

### Age

```
heart_attack |>
  group_by(glucose_level)|>
  summarise(Mean_age = mean(age),
            SD_age = sd(age),
            Min_age = min(age),
            Q1_age = quantile(age,0.25),
            Median_age = median(age),
            Q3_age = quantile(age,0.75),
            Max_age = max(age))
```

```
## # A tibble: 2 × 8
##   glucose_level Mean_age SD_age Min_age Q1_age Median_age Q3_age Max_age
##   <chr>            <dbl>  <dbl>   <dbl>  <dbl>      <dbl>  <dbl>   <dbl>
## 1 High              56.2   13.5      19     47         58   65       103
## 2 Normal            56.2   13.8      14     47         58   65.2     103
```

There appears to be little to no significant difference in age between the high glucose and normal glucose groups, based on the summary statistics calculated,This suggests that age may not be a significnat factor differentiating the two glucose categories, as the distributions for both groups are very similar

**Troponin**

```
heart_attack |>
  group_by(class)|>
  summarise(Mean_troponin = mean(troponin),
            SD_tropoin = sd(troponin),
            Min_troponin = min(troponin),
            Q1_troponin = quantile(troponin,0.25),
            Median_troponin = median(troponin),
            Q3_troponin = quantile(troponin,0.75),
            Max_troponin = max(troponin))
```

```
## # A tibble: 2 × 8
##   class    Mean_troponin SD_tropoin Min_troponin Q1_troponin
Median_troponin
##   <fct>            <dbl>      <dbl>        <dbl>       <dbl>
<dbl>
## 1 negative        0.0270      0.443        0.001       0.003
0.006
## 2 positive        0.571       1.39         0.003       0.016
0.044
## # ℹ 2 more variables: Q3_troponin <dbl>, Max_troponin <dbl>
```

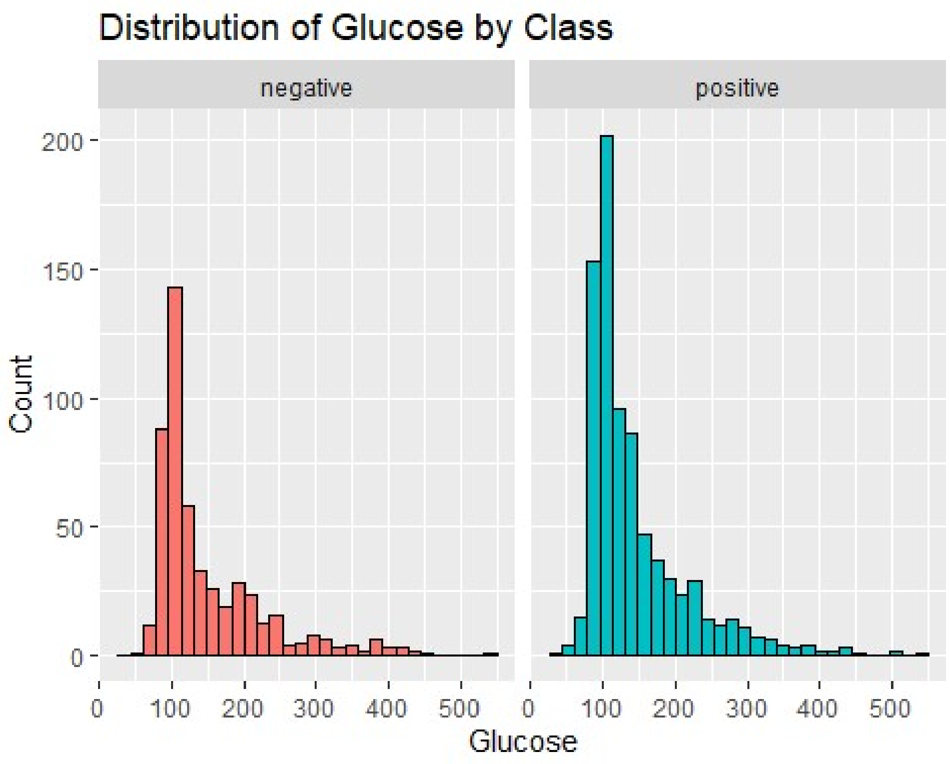
Troponin levels in the positive class show higher averages and greater variability, suggesting high levels of troponin are associated with a greater range of outcomes, possibly linked to heart attacks or severe heart conditions.
The negative class on the other hand has very low and consistent troponin levels, likely indicating healthier individuals with no significant heart damage.
The positive class has larger spread of values which further emphasizes the range of severity of the conditions associated with increased troponin levels.

**Visualizing Relationships Between Response and Explanatory Variables**

To gain deeper insights into how glucose levels vary across different heart attack outcomes, we generate graphs that visualize the relationship between multiple variables. Specifically, we will be creating a facet histogram that shows the distribution of glucose levels for positive (class = 1) and negative (class = 0) heart attack outcomes.

This approach allows us to observe how glucose levels differ between the two outcome classes, providing a clearer understanding of how this variable might influence the likelihood of a heart attack.

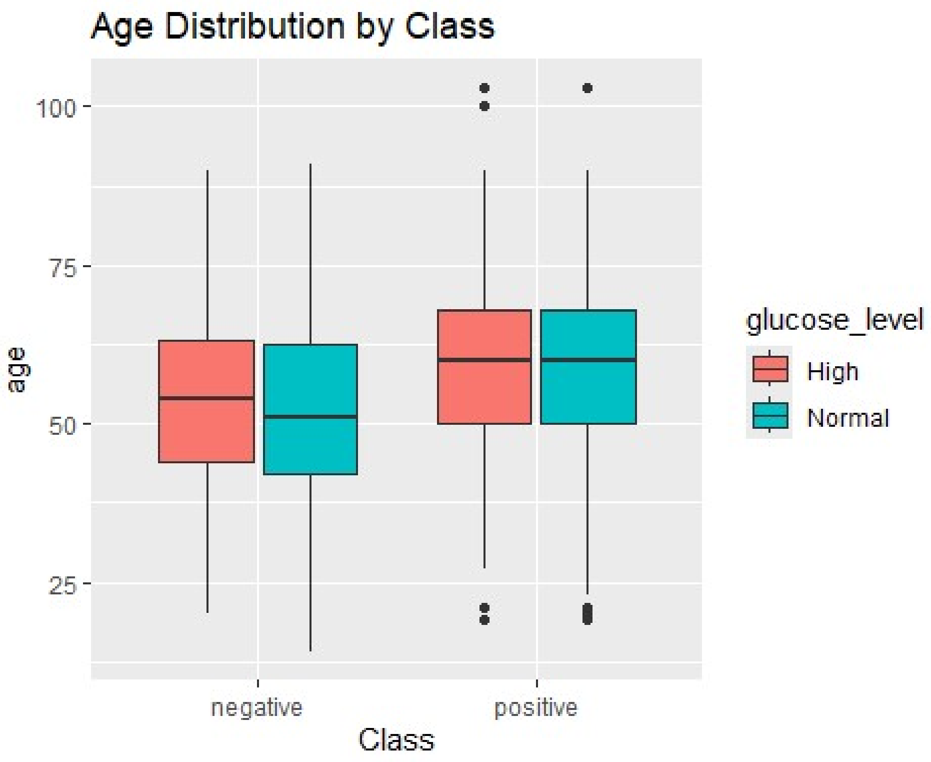
```
heart_attack |>
  ggplot(aes(x = glucose, fill = class)) +
  geom_histogram(bins = 30, color = "black",show.legend = F) +
  labs(title = "Distribution of Glucose by Class", x = "Glucose", y =
"Count")+
  facet_wrap(~class)
```



The histogram shows difference in glucose levels between classes,though this difference might not be very clear as expected,Patients with higher glucose levels appear more frequently in the positive class, suggesting that elevated glucose may be associated with heart attack risk.

Here we will use a boxplot to visualize how age distribution varies between heart attack outcome classes (positive vs. negative) while considering the influence of glucose levels (high vs. normal). This visualization helps us to explore whether age differs between patients with positive and negative outcomes and if the glucose level (high or normal) impacts the distribution of age.

```
heart_attack |>
  ggplot(aes(x= class,y = age,fill = glucose_level))+
  geom_boxplot()+
  labs(title = "Age Distribution by Class",x = "Class",y = "age")
```

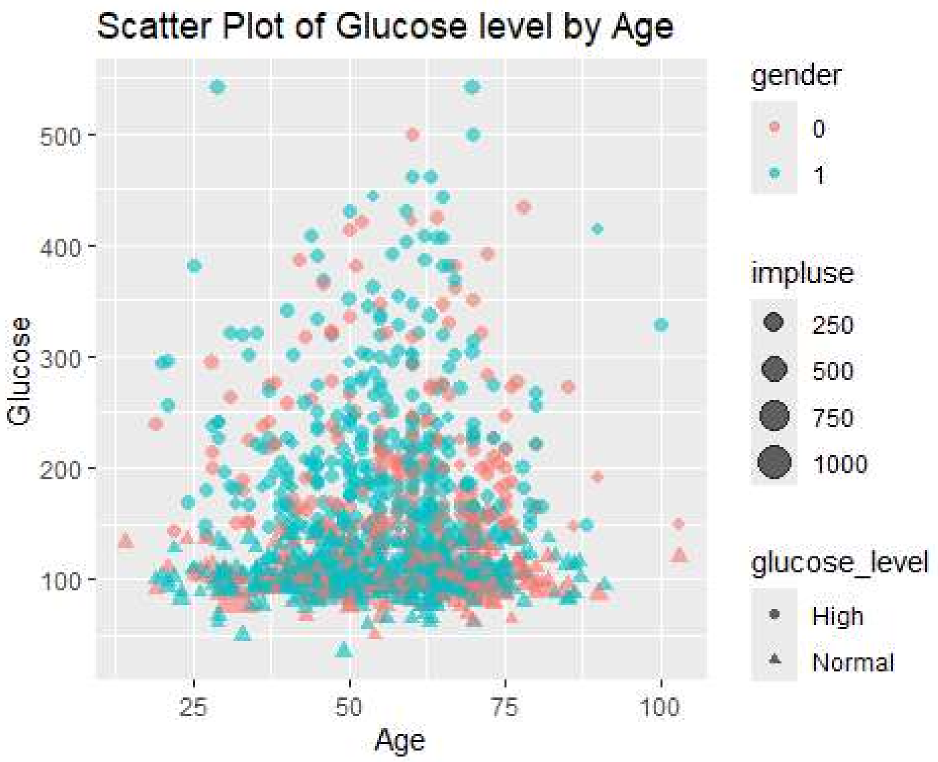


The age distribution by class thats negative and positive and glucose levels high and normal shows that age and glucose levels do not have clear relationship with class. Both classes display similar age distributions across high and normal glucose levels,suggesting that neither age nor glucose level alone effectively differentiates between the negative and positive classes in this dataset.

In this section, we will explore the relationship between glucose levels and age, while examining how these factors interact with gender, glucose levels (normal vs. high), and impulse. To achieve this, we use a scatter plot, that allows us to visualize the distribution of data and examine the associations between continuous variables, while also using categorical and quantitative features.

In the plot, the X axis represents age, while the Y axis displays glucose levels. The plot differentiates the data points by gender (using color), shows glucose level classification (normal or high) with distinct shapes, and uses the size of the points to reflect the impulse value, offering insights into its potential correlation with glucose and age

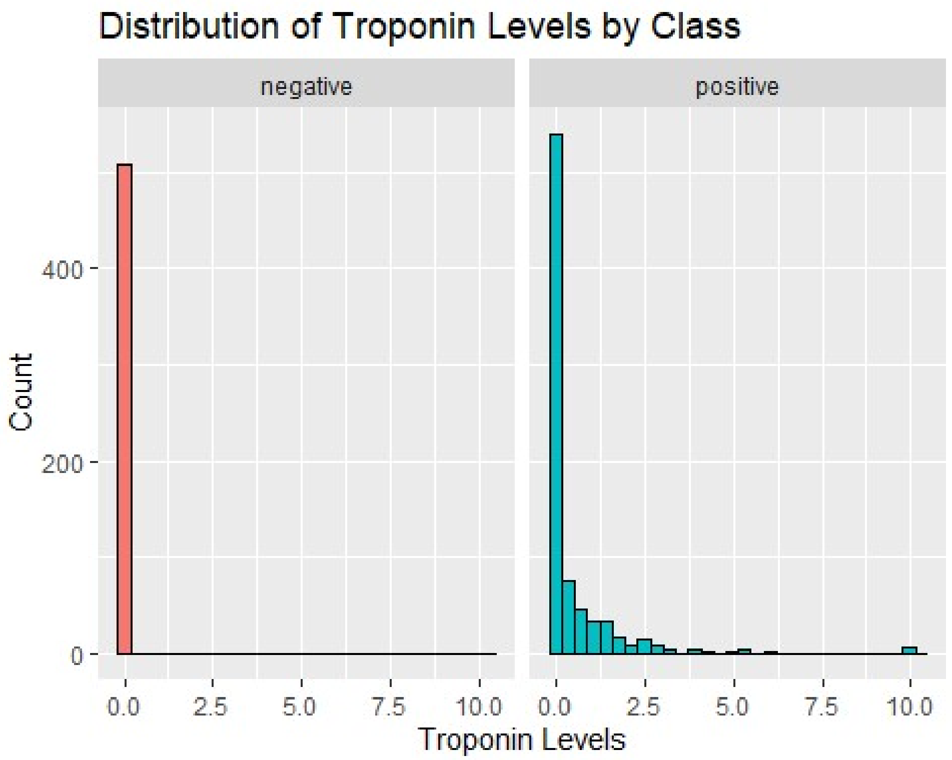
```r
heart_attack |>
  ggplot(aes(x = age,y = glucose,col = gender,shape = glucose_level,size =
impluse))+
  geom_point(alpha = 0.6)+
  labs(title = "Scatter Plot of Glucose level by Age",x = "Age",y = "Glucose"
)
```



The plot does not show a strong correlations between age, blood pressure, gender and impulse. This means that blood pressure and impulse may be influenced by other factors not shown in this plot, and that gender and glucose level do not significantly affect the age blood pressure relationship in this dataset.

We will use a histogram to examine the distribution of troponin levels across heart attack outcome classes (positive vs. negative outcomes). Troponin is a key to determine heart attack, and understanding its distribution can provide insights into the severity of heart attacks and its relationship with heart attack outcomes

```
heart_attack |>
  ggplot(aes(x = troponin, fill = class)) +
  geom_histogram(bins = 30, color = "black",show.legend = F) +
  labs(title = "Distribution of Troponin Levels by Class",x = "Troponin
Levels",y= "Count")+
  facet_wrap(~class)
```



The histogram plots show the distribution of troponin levels for two classes: negative and positive. Individuals in the negative class, troponin levels are more concentrated near zero. This indicates that lower troponin levels are strongly associated with the negative outcome, which means no heart attack.
The positive class has a more varied distribution of troponin levels, with a peak near zero but also a spread up to higher troponin values. This distribution suggests that higher troponin levels are associated with the positive outcome, possibly indicating heart attacks. The distribution of troponin levels is crucial for understanding the severity of heart attacks. Elevated levels of troponin are often associated with more severe heart muscle damage.

**Correlation**
In statistics, correlation refers to the statistical relationship between two variables, whether causal or not. It is used to determine how strongly two variables are related, with values ranging from −1 to 1. Here’s a breakdown of how correlation works:

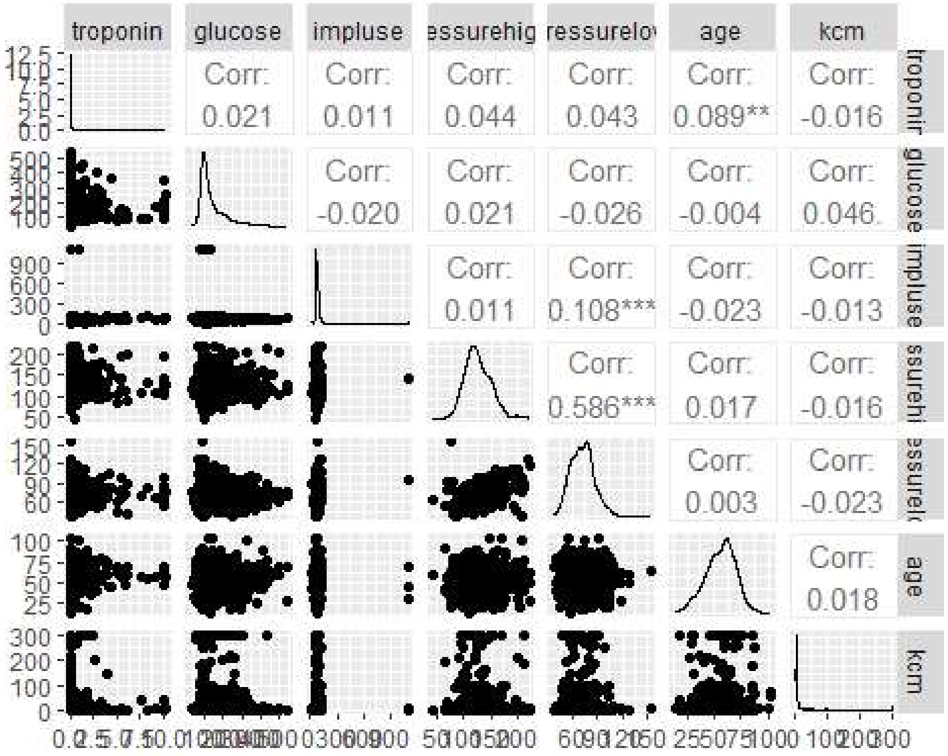
Positive Correlation (0 to 1): A positive correlation means that as one variable increases, the other variable also tends to increase. For instance, in a plot where both variables show an upward trend, their correlation would be positive. A correlation value closer to 1 suggests a strong positive relationship.

Negative Correlation (−1 to 0): A negative correlation indicates that as one variable increases, the other variable tends to decrease. A value closer to −1 suggests a strong inverse relationship.

No Correlation (around 0): If the correlation is near 0, it suggests no clear linear relationship between the variables.f

```
heart_attack %>%

dplyr::select(,troponin,glucose,impluse,pressurehight,pressurelow,age,kcm,)%>
%
  ggpairs()
```



We know that correlation valuesmeasures if two numberic values fits a linear relationship.From the plot those engineered features have very weak linear association with each other.

### Insights Based on Troponin as a Predictor

*Troponin* shows very weak linear relationships with the other variables in the dataset:
Correlation with *glucose* (0.021) and *impluse* (0.011) suggests minimal association.
Correlation with *pressure height* (0.044) and *pressure low* (0.043) implies weak relationships.
Correlation with *age* (0.089), though small, is statistically significant (indicated by **).
Correlation with *kcm* (-0.016) is negligible.
The significant correlation (0.586) between *pressure height* and *pressure low* indicates a strong relationship

### Modelling

In this section, I will employ Logistic Regression to analyze the relationship between the predictors and the categorical response variable (class).

```
heart_attack <- heart_attack |>
  mutate(class_01 = ifelse(class == "negative",0,1))
heart_attack
```

```
## # A tibble: 1,319 × 11
##      age gender impluse pressurehight pressurelow glucose    kcm troponin
class
##    <dbl> <fct>    <dbl>         <dbl>       <dbl>   <dbl>  <dbl>    <dbl>
<fct>
##   1    64 1            66           160          83     160   1.8      0.012
negat…
##   2    21 1            94            98          46     296   6.75     1.06
posit…
##   3    55 1            64           160          77     270   1.99     0.003
negat…
##   4    64 1            70           120          55     270  13.9      0.122
posit…
##   5    55 1            64           112          65     300   1.08     0.003
negat…
##   6    58 0            61           112          58      87   1.83     0.004
negat…
##   7    32 0            40           179          68     102   0.71     0.003
negat…
##   8    63 1            60           214          82      87 300        2.37
posit…
##   9    44 0            60           154          81     135   2.35     0.004
negat…
##  10    67 1            61           160          95     100   2.84     0.011
negat…
## # ℹ 1,309 more rows
## # ℹ 2 more variables: glucose_level <chr>, class_01 <dbl>
```

```
logistic_model <- glm(class_01 ~ troponin + glucose + impluse + age +
pressurehight + pressurelow + kcm, data = heart_attack, family = binomial)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(logistic_model)
```

```
##
## Call:
## glm(formula = class_01 ~ troponin + glucose + impluse + age +
##     pressurehight + pressurelow + kcm, family = binomial, data =
heart_attack)
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -3.8326878  0.5860257  -6.540 6.15e-11 ***
## troponin       5.6732140  0.7580101   7.484 7.19e-14 ***
## glucose       -0.0009123  0.0009598  -0.950    0.342    
## impluse        0.0003281  0.0015184   0.216    0.829    
## age            0.0485180  0.0058420   8.305  < 2e-16 ***
## pressurehight -0.0035143  0.0034944  -1.006    0.315    
## pressurelow    0.0027827  0.0063599   0.438    0.662    
## kcm            0.3599791  0.0363224   9.911  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 1759.2  on 1318  degrees of freedom
## Residual deviance: 1078.7  on 1311  degrees of freedom
## AIC: 1094.7
## 
## Number of Fisher Scoring iterations: 10
```

```
exp(coef(logistic_model))
```

```
##   (Intercept)      troponin       glucose       impluse           age 
##    0.02165134  290.96819422    0.99908811    1.00032817    1.04971430 
## pressurehight   pressurelow           kcm 
##    0.99649187    1.00278659    1.43329944
```

**Variables and Their Estimates**

The regression output provides the estimates for each predictor variable, which indicate their relationship with the likelihood of the outcome (class being positive).

**Intercept** (-3.8327): The baseline log-odds of the response variable (class_01 = 1, positive) when all predictor variables are at constant.

Odds Coefficient: $e^{-3.8327} = 0.0217$ (very low odds, meaning a very low probability of *positive* class at baseline).

**Troponin** (5.6732): A strong positive relationship between troponin levels and the likelihood of the class being "positive." A unit increase in troponin multiplies the odds by $e^{5.6732} = 290.97$ making troponin a very significant predictor.

*P-value* < 0.05 indicating strong evidence that this variable impacts the outcome

**Pressure High**(-0.00351): A weak negative relationship with the outcome. The odds decrease slightly by $e^{-0.00351} = 0.9965$ for each unit increase in pressure high.

*P-value*: 0.315, not statistically significant.

**Pressure Low** (0.00278): A weak positive relationship with the outcome. Each unit increase in pressure low slightly increases the odds by $e^{0.00278} = 1.0028$

*P-value*: 0.662, not statistically significant

*KCM* (0.36): A strong positive relationship with the likelihood of a positive class. Each unit increase in KCM multiplies the odds by $e^{0.36} = 1.4333$
meaning a 43.3% increase in odds.

*P-value* < 0.001 which makes KCM highly significant

**P-Values**
The p-values help assess the statistical significance of each variable
- Significant predictors (p-value < 0.05): Troponin, Age, and KCM.
- Non-significant predictors (p-value ≥ 0.05): Glucose, Impluse, Pressure Height, and Pressure Low.

**Conclusion and Interpretation of Findings**

The analysis conducted for the heart attack dataset successfully addressed the research questions posed at the start of the project. Below is a summary of the findings and their alignment with the questions:

**Research Question 1** What is the relationship between age, glucose levels, blood pressure, and troponin in patients, and how do they correlate with each other

*Findings*
Troponin levels emerged as a significant predictor, with a strong positive relationship to heart attack outcomes. Variables such as glucose levels, impulse, and blood pressure (both high and low) showed weak correlations with troponin and minimal association with each other. Correlation analysis revealed that only a few relationships, such as between pressure height and pressure low, were statistically significant

The results partially answers the research question. While some variables show relationships with heart attack outcomes, only troponin exhibits a substantial predictive capacity.

**Research Question 2**
Is age a significant in determining the likelihood of a positive or negative outcome

*Findings* The Logistic regression identified age as a significant factor, with older patients showing increased odds of a positive heart attack outcome.Age significantly influences heart attack outcomes, confirming the null hypothesis.

The project provides a clear understanding of the primary factors influencing heart attack outcomes. Troponin levels,age and KCM were identified as the most significant predictors, demonstrating strong associations with the likelihood of a positive outcome. However, other variables such as glucose, blood pressure, and impulse showed weaker correlations and limited predictive value in the model.

**REFERENCE**

**DR.Mostafa Sayed**, Ph.D.: Instructor of the course, who provided guidance and insights throughout the project.

**Oluwatobi Akinbode**: Teaching assistant, who offered additional support and helped clarify analytical concepts.

**Kaggle** : Source of the dataset used for the analysis. The dataset provided real-world data to explore predictive factors for heart attack outcomes.
https://www.kaggle.com/datasets/bharath011/heart-disease-classification-dataset/data

Rafael A. Irizarry (2019). Introduction to Data Science: Data Analysis and Prediction Algorithms with R. A foundational text used to understand statistical concepts and techniques applied in this project.